

University of Groningen

Nucleotide Sequence of the Cell Wall Proteinase Gene of *Streptococcus cremoris* Wg2

Kok, Jan; Leenhouts, Kees J.; Haandrikman, Alfred J.; Ledebor, Aat M.; Venema, Gerhardus

Published in:
Applied and environmental microbiology

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1988

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kok, J., Leenhouts, K. J., Haandrikman, A. J., Ledebor, A. M., & Venema, G. (1988). Nucleotide Sequence of the Cell Wall Proteinase Gene of *Streptococcus cremoris* Wg2. *Applied and environmental microbiology*, 54(1), 231-238.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Nucleotide Sequence of the Cell Wall Proteinase Gene of *Streptococcus cremoris* Wg2

JAN KOK,^{1*} KEES J. LEENHOUTS,¹ ALFRED J. HAANDRIKMAN,¹ AAT M. LEDEBOER,²
AND GERARD VENEMA¹

Institute of Genetics, University of Groningen, Kerklaan 30, 9751 NN Haren,¹ and Unilever Research Laboratorium, Olivier van Noortlaan 120, 3133 AT Vlaardingen,² The Netherlands

Received 26 June 1987/Accepted 22 October 1987

A 6.5-kilobase *Hind*III fragment that specifies the proteolytic activity of *Streptococcus cremoris* Wg2 was sequenced entirely. The nucleotide sequence revealed two open reading frames (ORFs), a small ORF1 with 295 codons and a large ORF2 containing 1,772 codons. For both ORFs, there was no stop codon on the *Hind*III fragment. A partially overlapping *Pst*I fragment was used to locate the translation stop of the large ORF2. The entire ORF2 contained 1,902 coding triplets, followed by an apparently rho-independent terminator sequence. The inferred amino acid sequence would result in a protein of 200 kilodaltons. Both ORFs have their putative transcription and translation signals in a 345-base-pair *Cla*I fragment. ORF2 is preceded by a promoter region containing a 15-base-pair complementary direct repeat. Both the truncated 33- and the 200-kilodalton proteins have a signal peptide-like N-terminal amino acid sequence. The protein specified by ORF2 contained regions of extensive homology with serine proteases of the subtilisin family. Specifically, amino acid sequences involved in the formation of the active site (viz., Asp-32, His-64, and Ser-221 of the subtilisins) are well conserved in the *S. cremoris* Wg2 proteinase. The homologous sequences are separated by nonhomologous regions which contain several inserts, most notably a sequence of approximately 200 amino acids between the His and Ser residues of the active site.

Because of their importance in the development of flavor and texture in a wide variety of fermented foods throughout the world, the proteolytic enzymes of lactic acid bacteria have been the subject of extensive research during the last decade. These studies have revealed the existence of an astonishingly complex system of proteinases and peptidases. Until now, attempts to unravel the complexity of the proteolytic systems have concentrated on the localization and biochemical characterization of these activities. Intracellular proteinases, as well as several different proteolytic activities associated with the cell wall, have been reported for *Streptococcus cremoris* and *S. lactis* (for reviews, see references 16 and 35). Exterkate (6) distinguished three proteolytic activities, on the basis of pH and temperature optima, which were present in different combinations in the cell walls of different *S. cremoris* strains. In *S. lactis*, multiple proteinases have been demonstrated by a zymogram staining technique on lysozyme-treated cells (3). In an attempt to ascribe proteolytic activities to separate enzymes, Hugenholtz et al. (12) have used antibodies against the purified proteolytic systems. Characteristic combinations of protein peaks in crossed immunoelectrophoresis experiments were the basis of a new classification of the proteolytic systems of *S. cremoris* strains. The overall impression of the biochemical data available is that the cell wall bound proteinases of lactic acid streptococci are very large enzymes (with molecular weights of 130,000 or more) which require Ca^{2+} ions for stabilization in an active configuration. Inhibition studies show that they are serine proteinases (8, 9, 13).

In a previous paper, we reported on the cloning and expression of the genetic information of the proteolytic activity of *S. cremoris* Wg2 (14). A 6.5-kilobase *Hind*III fragment from the proteinase plasmid pWV05 of this strain complemented the proteinase deficiency in *S. lactis* (Prt⁻). It

specified two proteins, A and B, of the proteolytic system of *S. cremoris* Wg2 in *S. lactis* as well as in *B. subtilis*. Here we report on the nucleotide sequence of the gene that specifies the cell wall-bound proteinase of *S. cremoris* Wg2 and on some characteristics of this sequence and of the enzyme as deduced from the predicted amino acid sequence.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. Growth and maintenance of bacteria and selective conditions have been previously described (14). Plasmid pGKV500 (14) and its derivatives were constructed and maintained in *B. subtilis* PSL1 (21). *Escherichia coli* JM101 (44) was used as the host for M13 and its derivatives.

Molecular cloning techniques. Plasmid DNA was isolated as described previously (14). Restriction nuclease enzymes, T4 DNA ligase, and the Klenow fragment of *E. coli* DNA polymerase I were purchased from Boehringer Mannheim Biochemicals and used as recommended by the manufacturer. General procedures for cloning and DNA manipulations were essentially as described by Maniatis et al. (18). Competent cells of *E. coli* were transformed as described by Mandel and Higa (17). Protoplasts of *B. subtilis* were transformed as described by Chang and Cohen (2).

DNA sequence analysis. Subfragments of the 6.5-kb *Hind*III fragment of pGKV500 were cloned in both orientations in phage M13 by using M13 mp10, mp11, mp18, and mp19 (44). The dideoxynucleotide sequencing method of Sanger et al. (25) was used with buffer gradient gels and [α -³⁵S]dATP (1). Synthetic 17-mer primers were prepared on a model 380A DNA synthesizer (Applied Biosystems, Foster City, Calif.) and purified on 20% polyacrylamide gels. To confirm the nucleotide sequence around the restriction enzyme sites used for cloning in M13, a sequence reaction was performed on pGKV500, which was made single stranded in the region of interest by cutting with an appropriate restric-

* Corresponding author.

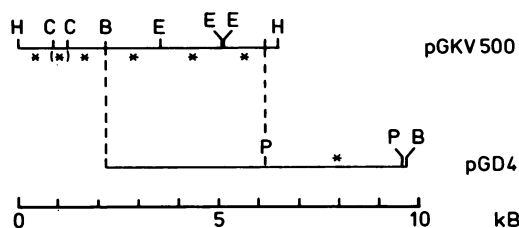


FIG. 1. Part of the *S. cremoris* Wg2 proteinase plasmid pWV05, which specifies cell wall-bound proteolytic activity. pGKV500 carries the 6.5-kb *Hind*III fragment which complements proteinase activity. pGD4 contains a partially overlapping 7.5-kb *Bam*HI fragment cloned in *E. coli* (14). The fragments cloned in M13 are indicated by an asterisk. Abbreviations: B, *Bam*HI, C, *Cla*I, E, *Eco*RI, H, *Hind*III; P, *Pst*I.

tion enzyme and subsequent treatment with *E. coli* exonuclease III (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) as advised by the manufacturer. Exonuclease III-treated DNA (1.5 to 2 μ g) was used in a standard sequencing reaction. Nucleotide sequences were stored, matched, and processed by using the computer programs of Staden (29–31).

RESULTS

M13 cloning and DNA sequencing. Figure 1 shows a 6.5-kb *Hind*III fragment of the proteinase-specifying plasmid pWV05 of *S. cremoris* Wg2 and an overlapping *Bam*HI fragment of 7.5 kb. The *Hind*III fragment, introduced in *S. lactis* (Prt[−]) on pGKV500, complemented the proteinase deficiency of this strain. The *Bam*HI fragment was cloned in pACYC184, resulting in pGD4 (14). All of the subfragments of the 6.5-kb *Hind*III fragment shown in Fig. 1 were cloned in both orientations in phage M13 mp10 and mp11 (44). Initially, the 345-base-pair (bp) *Cla*I fragment was found in a single mp10 clone, but it was lost upon subculturing. To determine the DNA sequence of the fragments, the two M13 clones of each fragment were sequenced in a cascade sequencing strategy with synthetic primers. After part of the sequence of a fragment had been determined by the dideoxynucleotide method (25), two primers were synthetically prepared. One primer was used to extend the nucleotide sequence, while a reversed primer was used to confirm the sequencing data by sequencing the opposite strand. To confirm the nucleotide sequence around the restriction enzyme sites used for subcloning in M13, pGKV500 was digested with an appropriate restriction enzyme and treated with exonuclease III to produce single-stranded DNA in the region of interest (26). This DNA was the template in a dideoxynucleotide sequencing reaction using one of the synthetic primers near the site to be sequenced. The exonuclease III strategy was also used to determine the nucleotide sequence of the 345-bp *Cla*I fragment. In this way, the nucleotide sequence of both strands of the entire *Hind*III fragment was obtained.

Codon preference analysis. Codon preference analysis (31) of the DNA sequence revealed two high-probability reading frames, one on each strand, orientated in opposite directions (data not shown). Both open reading frames (ORFs), one containing 295 codons (ORF1) and one with 1,772 codons (ORF2), had their endpoints outside the *Hind*III fragment. To extend the sequence of the large ORF2, a partially overlapping 3.5-kb *Pst*I fragment, isolated from pGD4 (Fig. 1), was cloned in M13 mp18 and mp19. With synthetic

primers, both strands of the left-hand part of this fragment were sequenced. A detailed restriction enzyme map deduced from this sequence and the position of the two ORFs are presented in Fig. 2. The *Hind*III fragment originally cloned in pGKV500 is shown shaded in gray. A third *Hind*III site, located 16 bp downstream of the second one, was not present in pGKV500. The first stop codon after ORF2 was found 380 bp downstream of the third *Hind*III site.

Nucleotide sequence. The nucleotide sequence of the *Hind*III fragment, extended with part of the DNA sequence of the *Pst*I fragment, is presented in Fig. 3. In Fig. 4, the 345-bp *Cla*I fragment containing the putative promoter region of both ORF1 and ORF2 is shown in more detail. ORF2 starts with an ATG start codon at position 1,206, and the first stop codon (TAG) is located at position 6,912, giving it a total length of 5,706 bp or 1,902 coding triplets. It has the potential to synthesize a protein of 200 kilodaltons (kDa). Upstream of the ATG start codon, around nucleotide 1,196, a Shine-Dalgarno sequence (GGAGG) similar to those reported for *B. subtilis* (11) is present, having a window of 10 bases and a free energy of binding of -14.4 kcal/mol (36). Although a second in-frame ATG codon is present at position 1,176, it is unlikely to be the start codon because it is not preceded by a reasonable ribosome-binding site. Starting at position 1,186, there are several potential promoter regions (20, 24). A continuous sequence of TTGAATTTGTTTC contains two putative -35 sequences. With a spacing of 16 and 15 bases, respectively, the two -35 sequences are followed by two overlapping consensus -10 regions (TATAATA-TAAT, starting at position 1,106). The region from position 1,106 to 1,141 contains several other partially overlapping Pribnow boxlike sequences. Upstream of the -35 region, there is an AT-rich region (86% AT over the first 50 bases), with several alternating stretches of A's and T's which resemble the signals known to enhance transcription in *B. subtilis* (4). Actually, the whole region between the two *Cla*I sites, 345 bp in length, is AT rich (73% A's and T's). In the promoter region (from position 1,084 to 1,145), two long direct repeats of 15 and 13 bases are present. The promoter region further contains a long complementary inverted repeat starting at position 1,104. A hypothetical stem-loop structure with a calculated free energy of -10.2 kcal/mol (36)

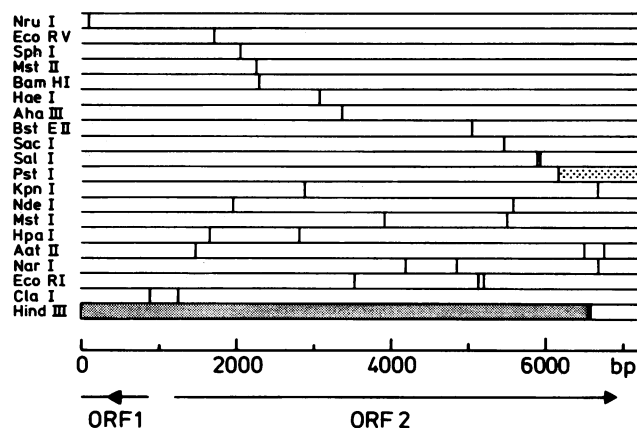


FIG. 2. Detailed restriction enzyme map of the proteinase region of pWV05 as deduced from the nucleotide sequence. The positions of ORF1 and ORF2 are indicated by the arrows. The *Hind*III fragment originally cloned in pGKV500 (14) is shaded, and the *Pst*I fragment used to extend the DNA sequence is stippled.

is depicted in Fig. 5A. The Pribnow box-rich region is completely buried in the proposed hairpin structure, thereby leaving the two -35 regions without their respective -10 regions.

The *Cla*I site at position 884 is located in the ATG start codon of ORF1 (Fig. 3). Nine bases upstream of this start codon, the sequence GAGGAGA constitutes a possible ribosome-binding site (11, 34). It is less clear-cut, however, to assign a promoter region upstream of this ribosome-binding site. There are several candidate -35 sequences, but only two of them have -10 regions which conform reasonably well to the consensus -10 sequence for *E. coli* and *B. subtilis* (20, 24). These are indicated by the leftward-directed arrows (at positions 1,155 and 1,132 and at positions 959 and 936) in Fig. 4. Promoter region 1,155/1,132 overlaps with the putative promoter for ORF2, and its -10 region is occupied in the stem of the proposed hairpin structure in this region (Fig. 5A).

The nucleotide sequence indicates that the codon usage in *S. cremoris* is quite different from that in *E. coli*. *S. cremoris* resembles *B. subtilis* in that it tends to distribute the codons for its amino acids more evenly (23).

Terminator structure downstream of ORF2. In the nucleotide sequence approximately 6,000 bp from the start of ORF2, a region of dyad symmetry is present between nucleotides 7,045 and 7,080, 130 nucleotides downstream of the TAG stop codon. It has all of the features of a rho-independent terminator sequence (24) consisting of two complementary inverted repeats which can form a stem of 15 bp (with seven G-C pairs and two mismatches). The hairpin structure is followed by a run of several T's and has a ΔG of -24.6 kcal/mol (36; Fig. 5B).

Putative signal peptides. The protein specified by ORF2 starts with a sequence of amino acids which closely resembles a typical signal peptide (39). Four positively charged amino acids are followed by a run of hydrophobic residues (Fig. 3). By the rules of von Heijne (40) for processing probability, a putative signal sequence cleavage site is situated between Ala-33 and Ala-34 in the canonical Ala-X-Ala-Ala sequence. Cleavage at this site would result in a signal peptide of 33 amino acids, which is in the size range reported for signal peptides of other gram-positive exoproteins. The 33-kDa protein coded for by ORF1 also contains a putative signal sequence structure with 32 amino acids.

Homology comparison. Of the proteins present in the National Biomedical Research Foundation protein data bank in October 1986, four showed homology with the *S. cremoris* Wg2 ORF2 protein. All four were bacterial serine proteases of the subtilisin family, and the overlaps are shown in Fig. 6. These subtilisins, produced by bacilli only, can be divided into two groups on the basis of structural and functional comparisons, including amino acid composition and sequence analysis, enzymatic activities, and immunological properties (22). Subtilisin Carlsberg and BPN' exemplify the two groups. At the amino acid sequence level, these two enzymes are approximately 70% homologous (22, 27). Obviously, ORF2 specifies a proteinase of the subtilisin type. One region of the *S. cremoris* proteinase, extending over 34 amino acids (amino acids 599 to 632), showed 50 to 56% homology with a region in the different subtilisins containing the reactive Ser-221. The corresponding serine in the *S. cremoris* proteinase, Ser-620, is contained in a stretch of seven amino acids with complete homology. A second region of homology, with 38 to 46% matches over a stretch of 117 amino acids, is found between amino acids 276 and 393 of the *S. cremoris* proteinase, corresponding with amino

acids 59 to 166 in the subtilisins. This region in the subtilisins includes the amino acids involved in the formation of the S1 specificity crevice (residues 125 to 127 and 152 to 154) and His-64, which, together with Ser-221 and Asp-32, constitutes the charge relay system crucial for enzyme activity (15). Asp-32 of the subtilisins is also found in a smaller region of homology with the *S. cremoris* proteinase. A stretch of seven amino acids around Asp-32, conserved in the subtilisins, is found around Asp-217 of the *S. cremoris* proteinase. In Fig. 6B, the results of the homology comparison are summarized and drawn to scale on a linear map. No homologies between the proteins present in the National Biomedical Research Foundation data bank and the truncated protein specified by ORF1 were found.

DISCUSSION

We sequenced over 7,000 bp of a region of the proteinase plasmid pWV05 of *S. cremoris* Wg2, which was shown to specify proteolytic activity. A 6,519-bp *Hind*III fragment contained within this sequence restored the proteolytic deficiency in *S. lactis* (Prt⁻) (14). From the two incomplete ORFs found on the fragment, only the largest was sequenced to its end. The first stop codon was located 5,706 bp downstream of the ATG start. The proposed transcription- and translation-regulatory sequences of this lactic acid streptococcal gene closely resemble those reported for *B. subtilis* and *E. coli* (11, 20, 24, 34) and are in good agreement with the sequences determined by van der Vossen et al. (36a). Metabolic regulation of proteinase synthesis in lactic acid streptococci has been observed (6, 7, 12, 16). The occurrence of a 36-bp region of dyad symmetry in the promoter region is suggestive of a regulatory region and might be a binding site for a regulatory protein (24). A similar region of dyad symmetry has been reported in front of the *sprE* gene that encodes the *B. subtilis* subtilisin E protease, a gene which is under catabolite repression (42).

From the homology comparison with the subtilisins, it is obvious that ORF2 specifies a serine protease. This finding is in accordance with the results of inhibition studies on the purified enzyme showing its sensitivity to the serine protease inhibitor phenylmethylsulfonyl fluoride (8, 9, 13). The three most-conserved regions include the triplet Arg-32, His-64, and Ser-221 of the reactive center of subtilisin (Arg-217, His-281, and Ser-620 in the *S. cremoris* proteinase). In the four enzymes compared, the three regions Asp-32/Asp-217, His-64/His-281, and Ser-221/Ser-620 share 50, 43, and 53% identical residues, respectively (Fig. 6A). When the percent match with at least one of the subtilisins was calculated and the most conservative amino acid replacements were regarded as identical residues (5), the homology increased to 83, 55, and 68%, respectively. The stretch of 107 amino acids around His-64 in the subtilisins also includes two sequences involved in the formation of the S1 specificity crevice (Ser-125-Leu-126-Gly-127 makes up one side of this pocket, and the side chains of Ala-152-Ala-153-Gly-154 form the other side [15]). The analogous sequences in the *S. cremoris* proteinase are Ser-349-Leu-350-Gly-351 and Ser-380-Ala-381-Gly-382. The latter region is part of a longer stretch of complete homology and includes the highly conserved Asn-155 (Asn-383 in the *S. cremoris* proteinase), which is important for stabilization of the reaction intermediate formed during proteolysis (15). The most striking feature of the *S. cremoris* proteinase in comparison with the subtilisins, however, is the presence of several stretches of amino acids not found in the subtilisin sequences (Fig. 6B). The distance

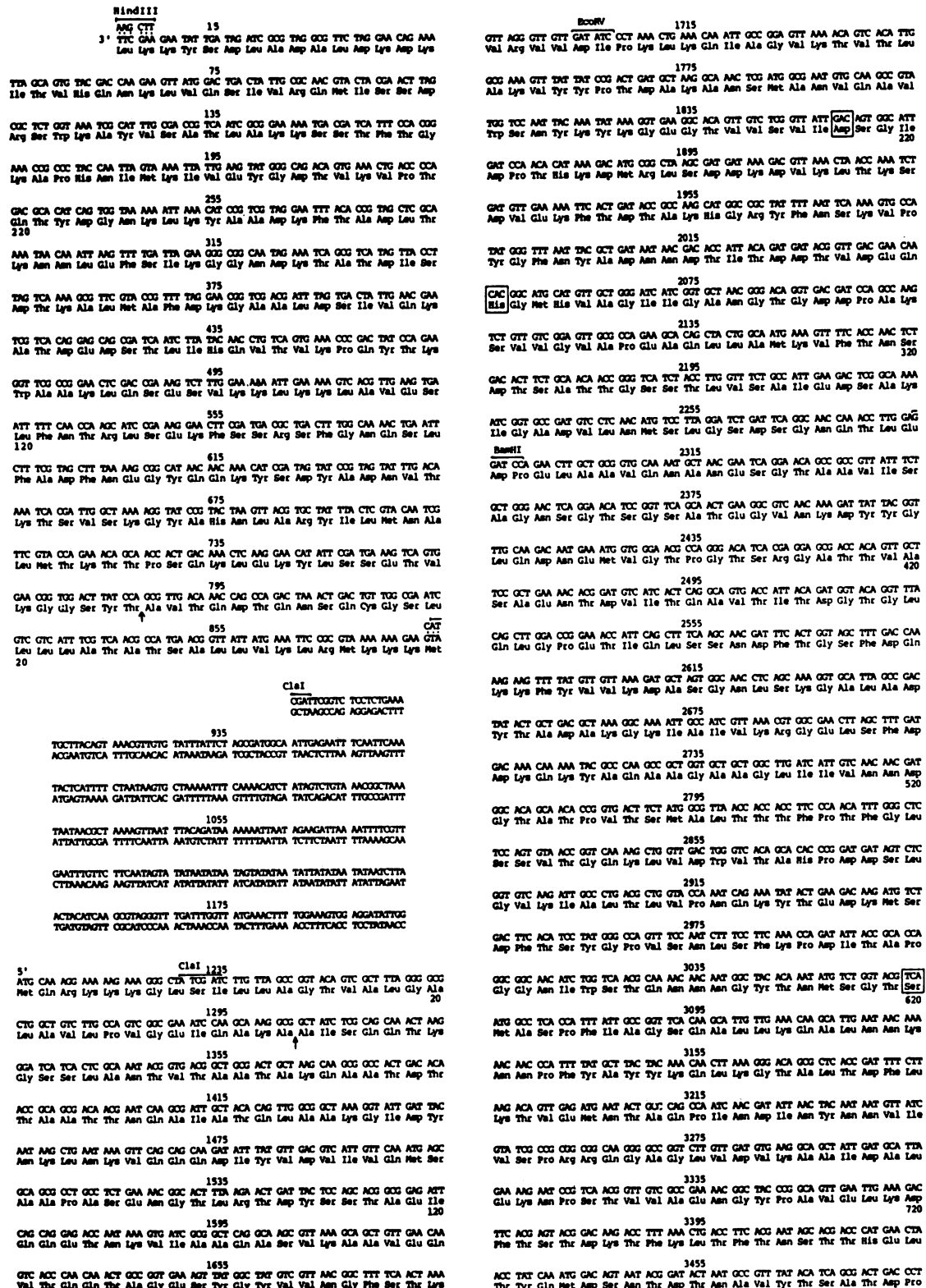


FIG. 3. Nucleotide sequence and inferred amino acid sequence of the *S. cremoris* Wg2 proteinase gene and its flanking regions. For both ORFs, the sequence of the nontranscribed DNA strand is presented. Numbering of the nucleotides is from the leftmost *HindIII* site. Amino acid numbering is shown under the sequence. The small untranslated region from position 886 to 1,205 contains the putative -35 and -10 sequences for ORF1 and ORF2 and is shown in more detail in Fig. 4. The putative signal sequence cleavage sites are shown by small vertical arrows. Asp-217, His-281, and Ser-620 are boxed. At the 3' end of the nucleotide sequence, the putative terminator of ORF2 is indicated. For details, see the text.

3515
 AAT TCT GGG GTT TTT TAT GAC AAG AAT ATT GAT GGA GCA GGC ATT AAA GCT GGC AGT AAC
 Aen Ser Gly Val Leu Tyr Asp Lys Lys Ile Asp Gly Ala Ala Ile Lys Ala Gly Ser Aen
 3575
 ATA ACT GTG CTT GCT GGG AAA AGC GCG CAG ATT GAA TTC ACA CTA TCT TTG CCG AAG TCT
 Ile Thr Val Pro Ala Gly Lys Thr Ala Gln Ile Glu Phe Thr Leu Ser Leu Pro Lys Ser
 3635
 TTT GAC CAA CAG CAA TTT GTT GAA GGT TTT CTG AAC TTT AAG GGT GCA GAT GCA TCG GGC
 Phe Asp Gln Gln Gln Phe Val Glu Gly Phe Leu Aen Phe Lys Gly Ser Asp Gly Ser Arg
 820
 3695
 TTG AAC TTG CCA TNC ATG GGC TTT TTT GGT GAC TCG AAT GAC GGT AAT GAT GTC GAT AGT
 Leu Aen Leu Pro Tyr Met Gly Phe Phe Gly Asp Trp Aen Asp Gly Lys Ile Val Asp Ser
 3755
 CTC AAT GGG ATC ACT TAT AGT GCT GGT GGT AAT TTT GGC ACC GTG CCA CTA TTG AGC
 Leu Aen Gly Ile Thr Tyr Ser Pro Ala Gly Gly Aen Phe Gly Thr Val Pro Leu Leu Thr
 3815
 AAC AAA AAT ACA GGC ACT CAA TAT TAT GGT GAT GTC ACA GCT GAT GGC AAC CAG
 Aen Lys Aen Thr Gly Thr Gln Tyr Tyr Gly Gly Met Val Thr Asp Ala Asp Gly Aen Gln
 3875
 ACA GTT GAC GAT CAG GCG ATT GCT TTT TCG AGT GAC AAG AAT GCT TTA TAT AAT GAC TAT
 Thr Val Asp Gln Gln Tyr Ile Tyr Tyr Aen Ala Pro Ala Trp Asp Gly Thr Tyr Tyr Aen
 3935
 AGC ATG AAG TAT TAT CTA TTG GGC AAT ATC ACC AAC GTC ACA GTT GAT ATT CTT GAT GGT
 Ser Met Lys Tyr Tyr Leu Leu Arg Aen Ile Ser Aen Val Gln Val Asp Ile Leu Asp Gly
 920
 3995
 CAG GGC AAT AAA GTT ACG ACT CTC AAT TCC ACC AAT CTG ACG AAG ACC TAT TAT AAT
 Gln Gly Aen Lys Val Thr Thr Leu Ser Ser Ser Thr Aen Leu Thr Lys Thr Tyr Tyr Aen
 4055
 GCT CAT TCG CAG CAG TAC ATC TNC TNC AAT GCT CCA GCG TCG GAT GGC ACC TAT TAT GAT
 Ala His Pro Ser Gln Gln Tyr Ile Tyr Tyr Aen Ala Pro Ala Trp Asp Gly Thr Tyr Tyr Aen
 4115
 CAA GGT GAT GGC AAC ATC AAG ACG GCT GAT GAT GGC AGT TAT TAT TAT GGT ATT TCC GGT
 Gln Arg Asp Gly Aen Ile Lys Thr Ala Asp Asp Gly Ser Tyr Thr Tyr Arg Ile Ser Gly
 4175
 GTA CCG GAA GGC GGC AAC AAT CAA GTT TTT GAT GTG GCT TTC AAC CTC GAC TCT AAT
 Val Pro Glu Gly Gly Asp Lys Arg Gln Val Phe Asp Val Pro Phe Lys Leu Asp Ser Lys
 4235
 GCG GCG ACA GTT GCT CAT GTC GCT TTG TCA GGC AAA ACG GAA AAT GCG AAA ACC CAG TAT
 Ala Pro Thr Val Arg His Val Ala Leu Ser Ala Lys Thr Glu Aen Gly Lys Thr Gln Tyr
 1020
 4295
 TAT TTG ACA GCT GAA GGC AAG GAT GAT GAT GGT GCT GAT GGC AAG AAG GCT AAT AAA
 Tyr Leu Thr Ala Glu Ala Lys Asp Asp Leu Ser Gly Leu Asp Ala Thr Lys Ser Val Lys
 4355
 ACT GCA ATT AAT GAA GTG ACG AAT CTT GAT GCT ACC TTT ACC GAT GCT GGC ACA ACG GCT
 Thr Ala Ile Aen Glu Val Thr Aen Leu Asp Ala Thr Phe Thr Asp Ala Gly Thr Thr Ala
 4415
 GAT GGT TCT ACC AAA ATT GAA ACG CAA TTT TCT GAT GAA CAG GGC CAA GCA CTT GGC AAT
 Asp Gly Tyr Thr Lys Ile Glu Thr Pro Leu Ser Asp Glu Gln Ala Glu Ala Leu Gly Aen
 4475
 GGC AAC AAT TCG GCT GAG CTG TNC TTG ACT GAT AAT GCA TCC AAT GGC ACT GAT CAA GAT
 Gly Asp Aen Ser Ala Glu Leu Tyr Leu Thr Asp Aen Ala Ser Aen Ala Thr Asp Gln Asp
 4535
 GGC ACC GTT CAG AAG CCG GCG TCT ACA TCG TTT GAT TTA ATT GTG AAC GGC GGC GGT AAT
 Ala Ser Val Gln Lys Pro Gly Ser Thr Ser Phe Asp Leu Ile Val Aen Gly Gly Ile
 1120
 4595
 CCA GAG AAC ATT TCA ACT ACC ACA ACC GGC TNC GAA GGC AAT ACT CAA GGT GGC GCG ACG
 Pro Asp Lys Ile Ser Ser Thr Thr Thr Gly Tyr Glu Ala Aen Thr Gln Gly Gly Gly Thr
 4655
 TAT AGC TTT AGT GCA ACG TAT CCA GCG GTT GAC GGT ACT TNC ACT AAT CCA CAA GCA
 Tyr Thr Phe Ser Gly Thr Tyr Pro Ala Ala Val Asp Gly Thr Tyr Thr Aen Ala Gln Gly
 4715
 AAG AAA CAT GAT TTG AAC ACA ACC TNC GAT GCT GCG ACT AAC AGT TTC ACT GGC TCA ATT
 Lys Lys His Asp Leu Aen Thr Thr Tyr Asp Ala Ala Thr Aen Ser Phe Thr Ala Ser Met
 4775
 CCG GTC ACC AAT GCT GAT TNC GGC GCG CAA GTG GAT CTA TAT GGC GAT AAG GCG CAT ACC
 Pro Val Thr Aen Ala Asp Tyr Ala Ala Gln Val Asp Leu Tyr Ala Asp Lys Ala His Thr
 4835
 CAG TTG CTT AAA CAT TTT GAC ACC AAA GTT GCA CTG ACG GCG GCA ACC TTT ACT GAT TTG
 Gln Leu Leu Lys His Phe Asp Thr Lys Val Arg Leu Thr Ala Pro Thr Phe Thr Asp Leu
 1220
 4895
 AAA TTC AAC AAT GGC TCG GAT CAG ACC TCT GAA GCG ACC ATC AAG GTT ACA GCG ACG GTT
 Lys Phe Aen Aen Gly Ser Asp Gln Thr Ser Glu Ala Thr Ile Lys Val Thr Gly Thr Val
 4955
 AGT GCT GAC ACC AAG ACA GTT AAT GTT GGC AAC ACC GTC ACA CTA CTT GAT CCA CAA CAT
 Ser Ala Asp Thr Lys Thr Val Aen Val Gly Asp Thr Val Ala Ala Leu Asp Ala Gln His
 5015
 CAC TTT AGT GTT GAT GTA CCG GTT AAT TAT GGT GAC AAT ACC ATC AAG GTG ACC GGC ACC
 His Phe Ser Val Asp Val Pro Val Aen Tyr Gly Asp Aen Thr Ile Lys Val Thr Ala Thr
 5075
 GAC GAA GAT GGC AAC ACC ACG ACG GAG CAA AAG AAT ACC TCG TCT TAT GAT CTT GAT
 Asp Glu Asp Gly Aen Thr Thr Thr Glu Gln Lys Thr Ile Thr Ser Ser Tyr Asp Asp
 5135
 ATG TTG AAT TCT GTG ACG TTC GAT CAA GGT GTG ACA TTT GGT GGC AAT GAA TTC AAT
 Met Leu Lys Aen Ser Val Thr Phe Asp Gln Gly Val Thr Phe Gly Ala Aen Glu Phe Aen
 1320
 5195
 GGC ACC TCG GCT AAG TTC TAT GAC GCT AAG ACC GCG ATT GCG AAG ATT ACT GGT AAG GTC
 Ala Thr Ser Ala Lys Phe Tyr Asp Pro Lys Thr Gly Ile Ala Thr Ile Thr Gly Lys Val
 5255
 AAG CAC CCA ACG ACA CCG TTG CAG GTT GAT GGT AAG CAA ATT CCA ATT AAG GAT GAT CTG
 Lys His Pro Thr Thr Thr Leu Gln Val Asp Gly Lys Gln Ile Pro Ile Lys Asp Asp Leu

5315
 ACT TTC AGT TTC ACT TTA GAT TTA GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
 Thr Phe Ser Phe Thr Leu Asp Leu Gly Thr Leu Gly Gln Lys Pro Phe Gly Val Val Val
 5375
 GGT GAC ACC ACT CAA AAC AAG ACC TTC CAA GAA CCG TTG ACC TTC ATT TTG GAT GCA GTG
 Gly Asp Thr Thr Gln Aen Lys Thr Phe Gln Glu Ala Leu Thr Phe Ile Leu Asp Ala Val
 5435
 GCT CCA ACA TTG TCA TTG GAG ACC TCG ACA GAT GCA CCG GTT TAT ACC AAT GAT CCA AAC
 Ala Pro Thr Leu Ser Leu Glu Ser Ser Thr Asp Ala Pro Val Tyr Thr Aen Asp Pro Aen
 1420
 5495
 TTC CAG ATT ACC GCA ACG GGC ACT GAC AAT GCG CAA TAT CTG ACT CTG CTA ATT AAC GGC
 Phe Gln Ile Thr Gly Thr Ala Thr Asp Aen Ala Gln Tyr Leu Ser Leu Ser Ile Aen Gly
 5555
 AGT TCT GTC GGC ACG CAA TAC GTA GAT ACT AAC ATC AAT AGT GGC AAA CCA GGT CAT ATG
 Ser Ser Val Ala Ser Gln Tyr Val Asp Ile Aen Ile Aen Ser Gly Lys Pro Gly His Met
 5615
 GCT ATT GAT CAG CCG GTT AAA TTG CTC GAA GGC AAA AAC GTG CTG ACT GGT GGT ACA
 Ala Ile Asp Gln Pro Val Lys Leu Leu Glu Gly Lys Aen Val Leu Thr Val Ala Val Thr
 5675
 GAT ACG GAA GAC AAC ACC ACG ACC AAG AAC ATC ACA GTT TAT TNC GAA CCA AAG AAA ACA
 Gly Ser Glu Asp Aen Thr Thr Thr Lys Aen Ile Thr Val Tyr Gly Leu Pro Gly Lys Thr
 5735
 CTG CCA CCA CCA ACT GTG ACG CCA ACT GAT GAA CCA GGC AAA GCG GTG ACT CTG ACG
 Leu Ala Ala Pro Thr Val Thr Pro Ser Thr Thr Glu Pro Ala Lys Thr Val Thr Leu Thr
 1520
 5795
 GCA AAC TCT GGC CCA ACG GGC GAA ACG GTT CAG TAT AGT GGT GAT GGT GGC AAG ACA TAT
 Ala Aen Ser Ala Ala Thr Gly Glu Thr Val Gln Tyr Ser Ala Asp Gly Gly Lys Thr Tyr
 5855
 CAG GAT GTT CCG GCA GGC GGT GTC ACC GTC ACG CAA AAT GGC ACC TTC AAG TTT AAG TCG
 Gln Asp Val Pro Ala Ala Gly Val Thr Val Thr Ala Aen Gly Thr Phe Lys Phe Lys Ser
 5915
 ACT GAT TTA TAT GGT AAT GAA TCA CCA GCG GTC GAT TAT GTT GTC ACC AAT AAT AAG GGC
 Thr Asp Leu Tyr Gly Aen Glu Ser Pro Ala Val Asp Tyr Val Val Thr Aen Ile Lys Ala
 5975
 GAT GAT GCT CCA CAA TTG CAG CCA GGT AAG CAG CAA CTG ACT AAT GCT ATT GCT TCC GGC
 Asp Asp Pro Ala Gln Leu Gln Ala Ala Lys Gln Glu Leu Thr Aen Leu Ile Ala Ser Ala
 6035
 AAA ACG CTA AGT GGC ACG GGT AAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
 Lys Thr Leu Ser Ala Ser Gly Lys Tyr Asp Asp Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr
 1620
 6095
 ACG CAG AAG CCA ACG GCG CTT GAT CAG ACG AAC GGC TCA GTT GAT TCA CTT ACT GGT
 Thr Gln Lys Ala Gln Thr Ala Leu Asp Gln Thr Aen Ala Ser Val Asp Ser Leu Thr Gly
 6155
 GGC AAT CCA CAG CTG CAA ACT GCG ATC AAT CAA TTA GCT GGC AAG TTA CTT GCA GAT AAG
 Ala Aen Arg Asp Leu Gln Thr Ala Ile Aen Gln Lys Leu Pro Ala Asp Lys
 6215
 AAG ACT TCG CTG CTT AAC CAG TTG CAA TCT GTG AAG GCT GGC CTG GCA ACG GAT TTG GGC
 Lys Thr Ser Leu Leu Aen Gln Leu Gln Ser Val Lys Ala Ala Leu Gly Thr Asp Leu Gly
 6275
 AAT CAA ACT GAT CCA ACG ACT GGC AAA ACA TTT ACG CCA GCG TTA GAC GAT CTA GTG CCA
 Aen Gln Thr Asp Pro Ser Thr Gly Lys Thr Phe Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 6335
 CAA GCT CAA CCA GCG ACG CAA ACG GAC CAG CAG CAT CAA GCG ACT CTT GCG AAG GTA CTT
 Gln Gln Gln Ala Gln Gly Thr Gln Thr Asp Asp Gln His Gln Thr Thr Thr Thr Thr Thr Thr Thr
 1720
 6395
 GAT CCA GTA TTA CAA AAT CTT GCG GAT GGT ATT AAA GCG CCA CCA GCG GGT GAT GGT GGT
 Asp Ala Val Leu Ala Lys Leu Ala Glu Gly Ile Lys Ala Ala Thr Pro Ala Glu Val Gly
 6455
 AAT GCT AAA GAT GCT CCA ACT GGC AAA ACT TCG TAT GGC GAT ATT GCT GAT CCA TTG ACG
 Aen Ala Lys Asp Ala Ala Thr Gly Lys Thr Trp Tyr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 6515
 TCT GGT CAA GCG AGT GGT GAT GCG TCT GAC AAG CTT CCA CAT TTA CAA GCT TTG CAA AGT
 Ser Gly Gln Ala Ser Ala Asp Ala Ser Asp Lys Leu Ala His Leu Gln Ala Leu Gln Ser
 6575
 CTG AAA ACG AAG GTG CCA GCT GGC GTT GAA GCG GGC AAC ACA GTT GGT AAA GGC GAC GGT
 Leu Lys Thr Lys Val Ala Ala Val Glu Ala Ala Lys Thr Val Gly Lys Gly Asp Gly
 6635
 ACA ACC GGT ACT GAC GAC AAA GCG GCG GGT CAA GGT ACC CCG GCG GCG GCT CCA GGC GAC
 Thr Thr Gly Thr Ser Asp Lys Gly Gly Gln Gly Thr Pro Ala Pro Ala Pro Gly Asp
 1820
 6695
 AAT GGT AAG GAC AAA GCG GAT GAG GCG ACG CAG CTT AGT TCT GGT AAT ATC CCA ACA
 Ile Gly Lys Asp Lys Gly Asp Glu Gly Ser Gln Pro Ser Ser Gly Gly Aen Ile Pro Thr
 6755
 AAT CCA GCG ACA ACG ACG TCA ACG ACG GAT GAT ACG ACT GAT GAT GAT GAT GAT GAT GAT
 Aen Pro Ala Thr Thr Thr Ser Thr Thr Asp Asp Thr Thr Asp Asp Aen Gly Gln Leu
 6815
 ACA TCC GGT AAG GCA CTA CCG AAG ACA GAG ACA ACT GAG CCG CCA GCG TTT GGT
 Thr Ser Gly Lys Gly Ala Leu Pro Lys Thr Gly Glu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 6875
 TTC TTG GGT GTC ATT GTG GTC ATT CTG AAG GGT TTA GCA TTG AAA CCG AAA CAA GGT
 Phe Leu Gly Val Ile Val Val Ile Leu Met Gly Val Leu Gly Leu Lys Arg Lys Gln Arg
 6935
 GAA GAA TNG OCTGTGACGT OCTTGGCTTT AGGCGACGCG OCTATGACCA AAGAGAGAG GGGCTGACCA
 Glu Glu ***
 6995
 TACTTGGAT TGAGGTTCAG AGTCANATG TACTGATGAC CCGTAAATTT TAAATATTTT
 7055
 AACCTTCTTT AGGAGCTCA CTTCACATCT TTGACATGCT CTGCTTTTCT CTGATGTTCT

between Asp-32 and His-64 in subtilisin is doubled to 64 amino acids, whereas His-64 and Ser-221 are spaced by an extra 182 amino acids in the *S. cremoris* proteinase (from 157 in subtilisin to 339 in the streptococcal proteinase). In the region where subtilisin has a small exterior loop (Gly-160 to Asn-163), the *S. cremoris* proteinase contains a stretch of approximately 180 amino acids not found in subtilisin. Because the spatial relationship among the amino acids of the active center, the S1 specificity crevice, and Asn-383 are kept intact, we may speculate that this large insert (and perhaps some of the smaller ones) can be envisaged as protruding from a subtilisinlike core.

The predicted amino acid sequence gives the *S. cremoris* proteinase a calculated molecular weight of 200,000. This value cannot be easily reconciled with the results of Hugenholz et al. (13). These investigators showed that the proteolytic system of *S. cremoris* Wg2 consists of two proteinases, A and B, with estimated molecular weights of 140,000 each, based on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Part of the difference in size can be explained by the assumption that the *S. cremoris* proteinase is synthesized as a proenzyme, as are the subtilisins (33, 41). At the N terminus, there is a signal peptide-like sequence of 33 amino acids. The tentative cleavage site is separated from Asp-217 by 184 residues. If we assume that 30 to 40 amino acids are required for proper folding of the mature enzyme at the N terminus (in subtilisin, this number is 32), approximately 130 to 140 residues would remain, which might constitute a pro-region (sizes of gram-positive pro-sequences range from 77 to about 194 amino acids (37). From 16 to 17 kDa could be split off in this way from the N terminus. Recently, the gene for the extracellular serine protease of *Serratia marcescens* was cloned and sequenced (43). The mature protease is formed by processing of a proenzyme at the N terminus, as well as at the C-terminal part. The mature enzyme contains 388 amino acids (size, 41 kDa). The C-terminal peptide split off contains another 637 residues, with an approximate size of 70 kDa. Similarly, the *S. cremoris* proteinase might be processed at the C terminus, and this,

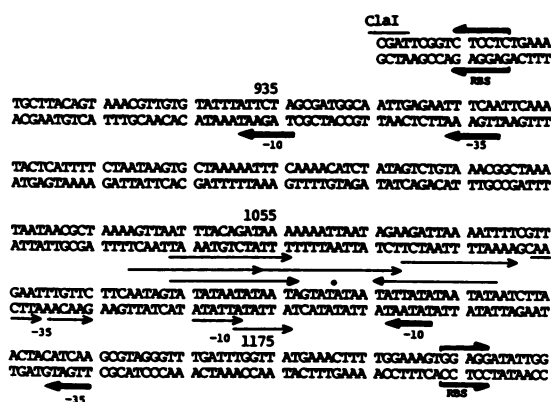


FIG. 4. Promoter region of ORF1 and ORF2. The 320-bp fragment shown (from position 886 to 1,205 in the nucleotide sequence of Fig. 3) contains the putative -35 and -10 regions for ORF1 (leftward-directed thick arrows under the sequence) and ORF2 (rightward-directed thin arrows under the sequence). Long arrows above the sequence, between position 1,099 and 1,141, indicate direct and inverted repeats. The two possible ribosome-binding site (RBS) sequences are indicated. The numbering of the nucleotides is the same as in Fig. 3.

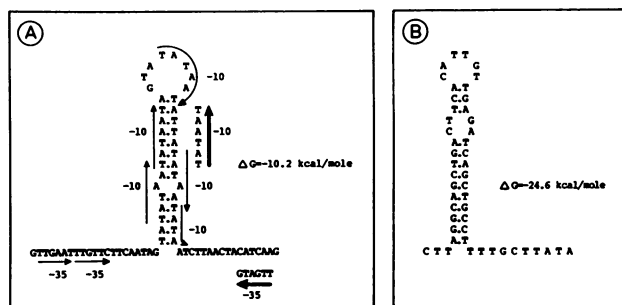


FIG. 5. Hypothetical stem-and-loop structures flanking the *S. cremoris* Wg2 proteinase gene. (A) Hairpin structure in the promoter region of ORF2. The -35 and -10 sequences of ORF2 are indicated by thin arrows. Part of the sequence is presented double stranded to show the possible promoter region, 1,155/1,132, of ORF1 (thick arrows). (B) Terminator structure 130 bases downstream of the TAG stop codon of ORF2.

together with the putative processing steps at the N terminus, might result in a mature enzyme of 140 kDa.

The observation that, upon prolonged incubation, the purified enzyme is subject to self-digestion might offer an alternative explanation. Because low (1 mM) concentrations of Ca^{2+} ions activate the similar *S. cremoris* AC1 proteinase (9), it is conceivable that, under the isolation conditions used, the streptococcal proteinase is released from the cell wall by a self-digestion step, resulting in the purification of a truncated protein of 140 kDa. Indeed, under certain conditions, proteinase activity can be isolated in protein bands with molecular weights as low as 60,000 (J. Erkelens, personal communication). Interestingly, at least one of these self-digestion sites, which have to be postulated to explain these observations, may be identified in the C terminus of the proteinase. The amino acid sequence Leu-1434 to Ser-1437 is identical to one of the digestion sites of the *S. cremoris* AC1 and *S. lactis* NCDO763 proteinase in β -casein (A. Geis and W. Bockelmann, personal communication; 19). This hypothesis is also in agreement with the genetic finding that the cloned *Hind*III fragment specifies a proteinase lacking 130 amino acids at the C terminus which still can complement proteinase deficiency. Moreover, a deletion in the gene removing the C-terminal 343 residues still specified an active enzyme (13a), showing that at least part of the C-terminal region can be deleted without severely affecting enzyme activity. This finding is in contrast with the situation for the extracellular proteolytic activity of *S. marcescens* in *E. coli*, which is lost upon introduction of frame shifts in the C-terminal part of the gene (43). Both processing and self-digestion might also offer an explanation for the localization of the genetic information for both proteins A and B (each with a size of 140 kDa) on the *Hind*III fragment in pGKV500 (14). Possibly, one of the proteins is a processing or breakdown product of the other which still exhibits proteinase activity. To match this with the crossed immunoelectrophoresis results, one would have to postulate the exposure of completely different antigenic determinants in A and B as a result of one of these digestion steps.

All *S. cremoris* proteinases are extremely specific and degrade only β -casein (9, 10, 19, 38), except the *S. cremoris* AM1 and SK11 proteinases, which also hydrolyze α -casein (38). A protein like bovine serum albumin, readily degraded by the subtilisins, is not hydrolyzed by *S. cremoris* proteinases (9). A differentiation between lactic acid streptococcal proteinases exists in the production of bitter peptides during

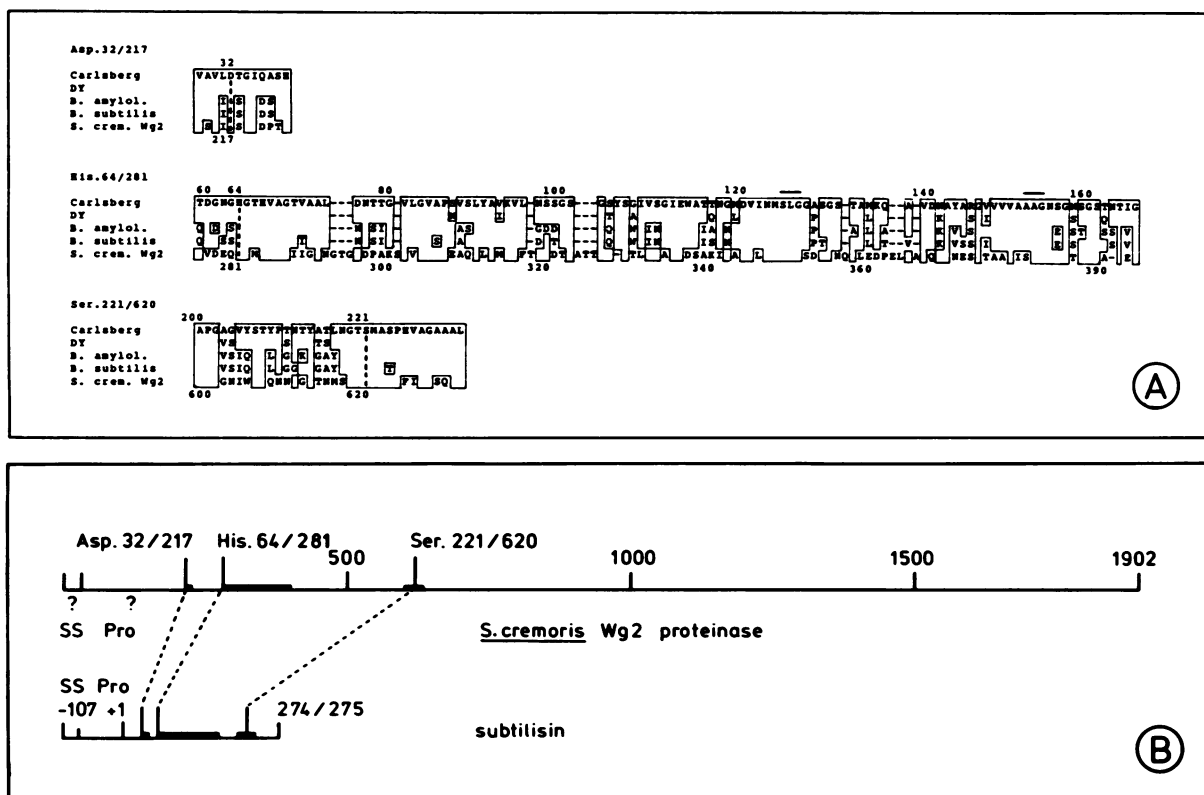


FIG. 6. Homology comparison. (A) Sequence homology of the *S. cremoris* Wg2 proteinase and subtilisins Carlsberg, DY, *B. amyloliquefaciens*, and *B. subtilis*. Sequences are from the National Biomedical Research Foundation protein data bank, October 1986. Only amino acids that differ from the residues in the Carlsberg enzyme are shown; identical residues are boxed. Asp, His, and Ser involved in the active site are indicated by vertical broken lines. The sequences forming the S1 specificity crevice are overlined. (B) The homologous regions from panel A (thick lines) were drawn to scale on a linear map of the whole proteinase and compared with a linear map of subtilisin. Numbers refer to amino acid residues. SS, Signal sequence; Pro, pro-sequence.

cheese production. This major flavor defect in cheese is thought to be related to proteinase (over) activity (16, 32). It will be interesting to learn whether the inserts or the long C terminus found in the *S. cremoris* Wg2 proteinase are involved in this specificity. Deletion analysis experiments are in progress to answer these questions and to find out whether the long C terminus plays a role in cell wall association, as suggested by the self-digestion hypothesis. We believe that the elucidation of the complete nucleotide sequence reported here is important in at least three respects. (i) It provides a basis for the construction of efficient expression and secretion vectors for lactic acid streptococci. (ii) It is essential for future research aimed to determine which parts of the enzyme are involved in its specificity. (iii) The nucleotide sequence is basic to research aimed at changing the properties of the enzyme to make it more suitable for dairying and, perhaps, other purposes.

ACKNOWLEDGMENTS

This work was supported by the Programme Committee on Biotechnology of the Netherlands and the Biomolecular Engineering Programme of the Commission of the European Communities.

We thank Beike Leegte for typing the manuscript and Henk Mulder for photography and preparation of the figures.

LITERATURE CITED

- Biggin, M. O., J. J. Gibson, and G. I. Hong. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**:3963-3965.
- Chang, S., and S. N. Cohen. 1979. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. *Mol. Gen. Genet.* **168**:111-115.
- Cliffe, A. J., and B. A. Law. 1985. Discontinuous polyacrylamide gel electrophoresis of cell wall proteinase from variants of *Streptococcus lactis*. *J. Appl. Bacteriol.* **58**:245-250.
- Doi, R. H. 1984. Genetic engineering in *Bacillus subtilis*. *Biotechnol. Genet. Eng. Rev.* **2**:121-155.
- Doolittle, R. F. 1979. Protein evolution, p. 2-118. In H. Neurath and R. L. Hill (ed.), *The proteins*, vol. 4. Academic Press, Inc., New York.
- Exterkate, F. A. 1976. Comparison of strains of *Streptococcus cremoris* for proteolytic activities associated with the cell wall. *Neth. Milk Dairy J.* **30**:95-105.
- Exterkate, F. A. 1985. A dual-directed control of cell wall proteinase production in *Streptococcus cremoris* AM1: a possible mechanism of regulation during growth in milk. *J. Dairy Sci.* **68**:562-571.
- Exterkate, F. A., and G. J. C. M. de Veer. 1987. Complexity of the native cell wall proteinase of *Lactococcus lactis* subsp. *cremoris* HP and purification of the enzyme. *Syst. Appl. Microbiol.* **9**:183-191.
- Geis, A., W. Bockelmann, and M. Teuber. 1985. Simultaneous extraction and purification of a cell wall-associated peptidase and β -casein specific protease from *Streptococcus cremoris* AC1. *Appl. Microbiol. Biotechnol.* **23**:79-84.
- Geis, A., B. Kiefer, and M. Teuber. 1986. Proteolytic activities of lactic acid streptococci isolated from dairy starter cultures. *Chem. Mikrobiol. Technol. Lebensm.* **10**:93-95.
- Hager, P. W., and J. C. Rabinowitz. 1985. Translational speci-

- ficity of *Bacillus subtilis*, p. 1–32. In D. A. Dubnau (ed.), *The molecular biology of the bacilli*. Academic Press, Inc., New York.
12. Hugenholtz, J., F. A. Exterkate, and W. N. Konings. 1984. The proteolytic systems of *Streptococcus cremoris*: an immunological analysis. *Appl. Environ. Microbiol.* **48**:1105–1110.
 13. Hugenholtz, J., D. van Sinderen, J. Kok, and W. N. Konings. 1987. The cell wall-associated proteases of *Streptococcus cremoris* Wg2. *Appl. Environ. Microbiol.* **53**:853–859.
 - 13a. Kok, J., D. Hill, A. J. Haandrinkman, M. J. B. de Reuver, H. Laan, and G. Venema. 1988. Deletion analysis of the proteinase gene of *Streptococcus cremoris* Wg2. *Appl. Environ. Microbiol.* **54**:239–244.
 14. Kok, J., J. M. van Dijk, J. M. B. M. van der Vossen, and G. Venema. 1985. Cloning and expression of a *Streptococcus cremoris* proteinase in *Bacillus subtilis* and *Streptococcus lactis*. *Appl. Environ. Microbiol.* **50**:94–101.
 15. Kraut, J. 1977. Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.* **46**:331–358.
 16. Law, B. A., and J. Kolstad. 1983. Proteolytic systems in lactic acid bacteria. *Antonie van Leeuwenhoek J. Microbiol. Serol.* **49**:225–245.
 17. Mandel, M., and A. Higa. 1970. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**:159–162.
 18. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 19. Monnet, V., D. Le Bars, and J. C. Gripon. 1986. Specificity of a cell wall proteinase from *Streptococcus lactis* NCDO763 towards bovine β -casein. *FEMS Microbiol. Lett.* **36**:127–131.
 20. Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* **186**:339–346.
 21. Ostroff, G. R., and J. J. Pène. 1983. Molecular cloning with bifunctional plasmid vectors in *Bacillus subtilis*: isolation of a spontaneous mutant of *Bacillus subtilis* with enhanced transformability for *Escherichia coli*-propagated chimeric plasmid DNA. *J. Bacteriol.* **156**:934–936.
 22. Ottesen, M., and I. Svendsen. 1970. The subtilisins. *Methods Enzymol.* **19**:199–215.
 23. Piggot, P. J., and J. A. Hoch. 1985. Revised genetic linkage map of *Bacillus subtilis*. *Microbiol. Rev.* **49**:158–179.
 24. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**:319–353.
 25. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
 26. Smith, A. J. H. 1979. The use of exonuclease III for preparing single stranded DNA for use as a template in the chain terminator sequencing method. *Nucleic Acids Res.* **7**:831–848.
 27. Smith, E. S., R. J. De Lange, W. H. Evans, M. Landon, and F. S. Markland. 1968. Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'; evolutionary relationships. *J. Biol. Chem.* **243**:2184–2191.
 28. Staden, R. 1984. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res.* **12**:505–519.
 29. Staden, R. 1984. Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Res.* **12**:521–538.
 30. Staden, R. 1984. Measurements of the effects that coding for a protein has on a DNA sequence and their use for finding genes. *Nucleic Acids Res.* **12**:551–567.
 31. Staden, R., and A. D. McLachlan. 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Res.* **10**:141–156.
 32. Stadhouders, J., G. Hup, F. A. Exterkate, and S. Visser. 1983. Bitter flavour in cheese. I. Mechanism of the formation of the bitter flavour defect in cheese. *Neth. Milk Dairy J.* **37**:157–167.
 33. Stahl, M. L., and E. Ferrari. 1984. Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro-derived deletion mutation. *J. Bacteriol.* **158**:411–418.
 34. Stormo, G. D., T. D. Schneider, and L. M. Gold. 1982. Characterization of translational initiation sites in *E. coli*. *Proc. Natl. Acad. Sci. USA* **81**:6115–6119.
 35. Thomas, T. D., and O. E. Mills. 1981. Proteolytic enzymes of starter bacteria. *Neth. Milk Dairy J.* **35**:255–273.
 36. Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Graller. 1973. Improved estimation of secondary structure in ribonucleic acids. *Nature (London) New Biol.* **246**:40–41.
 - 36a. van der Vossen, J. M. B. M., D. van der Lelie, and G. Venema. 1987. Isolation and characterization of *Streptococcus cremoris* Wg2-specific promoters. *Appl. Environ. Microbiol.* **53**:2452–2457.
 37. Vasantha, N., L. D. Thompson, C. Rhodes, C. Banner, J. Nagle, and D. Filpula. 1984. Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain a large open reading frame between the region coding for signal sequence and mature protein. *J. Bacteriol.* **159**:811–819.
 38. Visser, S., F. A. Exterkate, C. J. Slangen, and G. J. C. M. de Veer. 1986. Comparative study of action of cell wall proteinases from various strains of *Streptococcus cremoris* on bovine α_1 -, β -, and κ -casein. *Appl. Environ. Microbiol.* **52**:1162–1166.
 39. von Heijne, G. 1982. Signal sequences are not uniformly hydrophobic. *J. Mol. Biol.* **159**:537–541.
 40. von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133**:17–21.
 41. Wells, J. A., E. Ferrari, D. J. Henner, D. A. Estell, and E. Y. Chen. 1983. Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. *Nucleic Acids Res.* **22**:7911–7925.
 42. Wong, S.-L., C. W. Price, D. S. Goldfarb, and R. H. Doi. 1984. The subtilisin E gene of *Bacillus subtilis* is transcribed from a σ^{37} promoter in vivo. *Proc. Natl. Acad. Sci. USA* **81**:1184–1188.
 43. Yanagida, N., T. Uozumi, and T. Beppu. 1986. Specific excretion of *Serratia marcescens* protease through the outer membrane of *Escherichia coli*. *J. Bacteriol.* **166**:937–944.
 44. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. *Gene* **33**:103–119.